

ARTICLE

# Is FrameNet usage-based? Predicting frame element coreness with information theory and gradient boosting

Vladimir Buskin 

English Language and Linguistics, Katholische Universität Eichstätt-Ingolstadt, Germany  
Email: [vbuskin@ku.de](mailto:vbuskin@ku.de)

(Received 26 August 2025; Revised 27 January 2026; Accepted 16 March 2026)

## Abstract

The FrameNet project is a large-scale frame-semantic database with a seemingly usage-based core: It draws on 200,000 annotated sentences from representative corpora and offers the most comprehensive description of semantic valency patterns in English to date. Nevertheless, its empirical validity is weakened by the lack of statistical information on the distribution of lexical units, frames and frame elements. Similarly, the characterisation of frame elements as core, core-unexpressed, peripheral or extra-thematic – intended to indicate their essentiality to a frame – is primarily motivated on theoretical grounds. This raises the question of whether these labels are consistent with actual language use. After exhaustively extracting frequency data from Python’s NLTK FrameNet Corpus for all attested combinations of verbs, frames and frame elements, hierarchical gradient boosting models were trained on information-theoretic measures and word embeddings to predict the coreness of frame elements. The models provide strong usage-based evidence for a general core versus non-core distinction but cast doubt on further subdivisions such as core versus core-unexpressed or peripheral versus extra-thematic. While further validation is necessary, this contribution offers the first statistical perspective on the current state of FrameNet and its compatibility with usage-based approaches.

**Keywords:** argument structure; entropy; FrameNet; frame semantics; machine learning

## 1. Introduction

Frame semantic approaches to semantic and syntactic valency have not only been continuously refined since their inception but also increasingly applied to languages other than English (Boas, 2020; Boas et al., 2024; Fillmore et al., 2012). This is also true of the English FrameNet database,<sup>1</sup> which comprises over 13,000 lexical units in more

<sup>1</sup>Cf. <https://framenet.icsi.berkeley.edu> [Last accessed: June 03, 2025].



than 1,000 semantic frames with approximately 200,000 annotated sentences (Ruppenhofer et al., 2016, p. 7). In the broadest sense, the notion of ‘frames’ signifies ‘schematic representations of the conceptual structures and patterns of beliefs, practices, institutions, images, etc. that provide a foundation for meaningful interaction in a given speech community’ (Fillmore et al., 2003, p. 235). These frames group lexical units based on the frame elements, that is, the thematic (or  $\theta$ ) roles, they evoke (Ruppenhofer et al., 2016, pp. 7–8).

For instance, verbal lexical units such as *boil*, *brown*, or *fry* are said to evoke the Apply\_Heat frame,<sup>2</sup> in which a COOK applies heat with a certain TEMPERATURE\_SETTING to FOOD using a HEATING\_INSTRUMENT or a CONTAINER (cf. (1)–(3)).<sup>3</sup> While these frame element types are mostly specific to the APPLY\_HEAT frame, speakers may also choose to supply circumstantial elements, such as MANNER or PLACE, to flexibly modify this or other events. Frame elements are thus assumed to vary with respect to how essential they are to the frame they instantiate: While a cooking event is definitely conceivable without reference to a particular PLACE, this is not possible without a COOK or FOOD.

- (1) [*They*<sub>COOK</sub>] **boil** [*them*<sub>FOOD</sub>] [*in an iron saucepan*<sub>CONTAINER</sub>].
- (2) [*You*<sub>COOK</sub>] **can brown** [*it*<sub>FOOD</sub>] [*in hot fat*<sub>MEDIUM</sub>] [...].
- (3) [*She*<sub>COOK</sub>] **was frying** [*eggs and bacon and mushrooms*<sub>FOOD</sub>] [*on a camp stove*<sub>HEATING\_INSTRUMENT</sub>] [*in Woolley’s billet*<sub>PLACE</sub>].

FrameNet distinguishes between different degrees of ‘coreness’ among the elements evoked as part of a frame. More specifically, there are core and non-core elements, with the core category subsuming core and core-unexpressed elements, and the non-core category peripheral and extra-thematic ones. For example, COOK, FOOD, HEATING\_INSTRUMENT and CONTAINER are classified as core to the APPLY\_HEAT frame, whereas MEDIUM and PLACE are non-core. The non-core element PLACE is considered to belong to the subset of peripheral elements, which ‘do not introduce additional, independent or distinct events’ (Ruppenhofer et al., 2016, p. 24), while MEDIUM does, thus rendering it extra-thematic. As expected, core-unexpressed elements are rarely attested in FrameNet. The few existing examples list the ACT role, which is said to be evoked as part of the INTENTIONALLY\_ACT frame, as a core-unexpressed element (cf. (4)–(5)).

- (4) [*Iran*<sub>AGENT</sub>]’s **ability to conduct** [*these attacks*<sub>ACT</sub>] **was primarily the result of two factors.**
- (5) [...] **that** [*Iran*<sub>AGENT</sub>] **acted** [*in bad faith*<sub>MANNER</sub>] [*by failing to fully comply with IAEA inspections*<sub>ACT</sub>].

This raises the question of the grounds on which these coreness distinctions are made.

The FrameNet team classifies an element as core if it is (a) realised without exception, (b) assumes a definite reading upon omission, or is (c) formally or

<sup>2</sup>Cf. Ruppenhofer et al. (2016), p. 8; for all frame patterns and example sentences cited, see [https://franenet.icsi.berkeley.edu/fnReports/data/frameIndex.xml?frame=Apply\\_heat](https://franenet.icsi.berkeley.edu/fnReports/data/frameIndex.xml?frame=Apply_heat) [Last accessed: April 07, 2025].

<sup>3</sup>Frames and frame elements will be highlighted by small caps.

semantically idiosyncratic (Ruppenhofer et al., 2016, pp. 23–24). Core-unexpressed elements are claimed to be inherited from more abstract frames, and they are not available for independent realisation, as they are ‘absorbed by the lexical units in the frame’ (Ruppenhofer et al., 2016, p. 25). Of these criteria, only (a) can be directly quantified, suggesting that prototypical core elements occur with a probability equal to 1 or, at the very least, with a probability greater than that of core-unexpressed and non-core elements.<sup>4,5</sup> However, since FrameNet does not list ‘any statistical information about frequency of occurrence of syntactic patterns or about LUs [lexical units]’ (Fillmore et al., 2003, p. 248), this undermines the database’s reliability from a usage-based perspective. Similarly, in light of the quantitative turn of cognitive approaches to language (Janda, 2013), the lack of empirical evidence for FrameNet’s coreness labels could potentially undermine theoretical work relying on them.

In particular, research on argument omission would greatly benefit from the empirical validation of FrameNet’s coreness labels. In cognitive linguistics, omission phenomena are subsumed under the notion of Null Instantiation, whose scope only extends to core frame elements (Buskin, 2025, p. 284; Willich, 2022, p. 24; Ruppenhofer & Michaelis, 2014, p. 60; Fillmore, 1986). For instance, the FrameNet team deems CONTAINER a core element of the APPLY\_HEAT frame (cf. (1) earlier); therefore, this thematic role would be considered null-instantiated in (2) and (3), respectively; yet, this would not apply to the putative non-core elements MEDIUM and PLACE. Fundamentally, this issue boils down to what thematic roles can be considered conceptually obligatory for parsing a given event and when this obligatory status is attained.<sup>6</sup> Arriving at both a usage-based and reproducible method of capturing argument coreness would provide a robust baseline not only for further argument omission research (e.g., Chaves et al., 2025; Gavrusseva, 2024) but also for the cross-linguistic study of argument realisation more generally (e.g., Goldberg, 2005).

Previous quantitative research on the core/non-core distinction specifically has been conducted within the context of natural language processing (Matsubayashi et al., 2010; Nikolaev & Padó, 2023). Nikolaev and Padó (2023) have found that bidirectional encoder representations from transformers (BERT) are capable of distinguishing between core and non-core arguments, displaying substantial frequency effects (Nikolaev & Padó, 2023, p. 235). Yet, the study does not examine the core/non-core contrast itself and does not adduce pertinent accuracy metrics; instead, the emphasis lies on frame identification accuracy. Matsubayashi et al. (2010) focus on the classification of semantic roles by developing a new softmax classifier that assigns the most plausible frame element label to a constituent, given its frame and formal realisation (Matsubayashi et al., 2010, pp. 850–851). Their data indicate that core and non-core elements can be identified with very high accuracy

<sup>4</sup>This is in line with Jurafsky (1996), p. 167: ‘Obligatory arguments will have unity or near-unity probabilities, while optional arguments will have lower probabilities’.

<sup>5</sup>FrameNet’s core/non-core distinction is strikingly reminiscent of the argument/adjunct (or obligatory/optional) dichotomy in more traditional valency-based approaches to syntax (e.g., Herbst et al., 2004; Herbst & Götz-Votteler, 2007), which itself is strongly inspired by Lucien Tesnière’s *actants* and *circonstants*, as Willich (2022), p. 21 highlights (see also Tesnière et al., 2015, pp. Chapter 50).

<sup>6</sup>Scholars have taken issue with the binary ‘obligatory – optional’ distinction for decades. Despite attempting to develop more nuanced classifications, they ultimately rely on introspective methods (Ágel, 2000; Heidolph et al., 1984; Helbig, 1992; Vater, 1978).

using frame element descriptors such as SELLER (Matsubayashi et al., 2010, pp. 858–859), which were coined by the FrameNet annotators. Since the objective is to perform accurate semantic role labelling, the plausibility of the coreness labels is inconsequential.

This article addresses two issues: First, it aims to alleviate FrameNet's empirical opacity by supplementing it with statistical information on the attested combinations of verbs, frames and frame elements. Second, the article examines FrameNet's empirical plausibility by estimating how well usage data, which is derived from FrameNet's very own corpus examples, can explain the coreness distinctions made by its annotators. Since 'usage' can be conceptualised in different ways, it will be jointly captured in terms of information-theoretic measures and word embeddings, thereby catering to the probabilistic and distributional semantic characteristics of frame elements. Based on this data set, three supervised machine learning models are trained on the usage profiles of all retrieved frame elements to predict the following three coreness distinctions: (i) core versus non-core, (ii) core versus core-unexpressed and (iii) peripheral versus extra-thematic, respectively.

If the usage profiles meaningfully contribute to distinguishing between the different coreness categories, then the models should display high predictive accuracy. I hypothesise that the accuracy will be highest for (expressed) core elements, as this is the only element type for which FrameNet provides an overtly usage-based annotation criterion. Conversely, the remaining coreness types are expected to be less predictable, potentially resulting in poor model performance and hence less grounding in actual language use.

The analysis proceeds as follows: Section 2 formalises frame semantic concepts and their quantification, introduces information-theoretic measures and word embeddings and describes the statistical modelling approach. Section 3 presents results and model evaluation. Section 4 summarises the findings and discusses implications for frame semantics and cognitive linguistics.

## 2. Methodology

### 2.1. Formalisation and quantification

The FrameNet database, and hence the derived data set, is characterised by many-to-many relationships between verbs, frames and frame elements. In order to facilitate the statistical analysis, these interrelated sets and relations are formally modelled using set theory. Let  $V$  denote the set of verbs  $\{v_1, v_2, \dots, v_k\}$ , which serve as frame-evoking elements, and  $F$  the set of all indexed semantic frames  $\{f_1, f_2, \dots, f_j\}$ . Furthermore, let  $E = \{e_1, e_2, \dots, e_t\}$  denote the set of all distinct frame elements. These sets are jointly illustrated in Figure 1.<sup>7</sup> For instance, the lexical unit *leave*, depending on the context of its occurrence, may evoke the DEPARTING, GIVING, or the ABANDONMENT frame, among others. The participant roles associated with these frames often recur in other frames as well; the THEME argument is common to all three of them, whereas DONOR and AGENT are not. To avoid ambiguity, frame elements should always be characterised with respect to their underlying frame and the evoking verb.

<sup>7</sup>Note that the verbs evoke more frames and frame elements than shown in this figure. For a complete account, see the respective lexical entries in FrameNet.

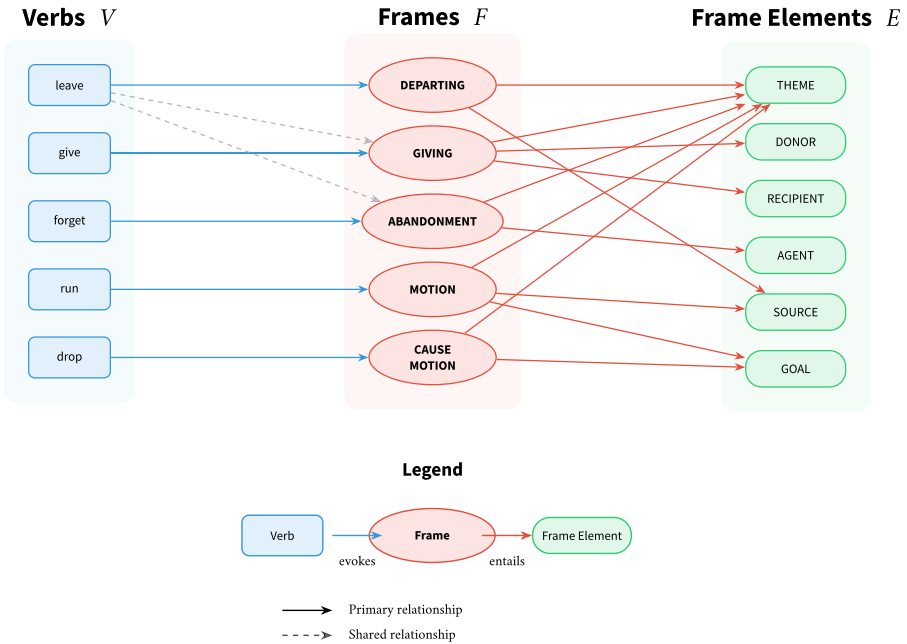


Figure 1. Simplified representation of set relations in the FrameNet database.

Consequently, this calls for relations that handle the evocation of frame elements and return the frame structure. To this end, define  $R_1 \subseteq V \times F$  and  $R_2 \subseteq F \times E$ , so that the set of elements conditioned on a specific verb and a specific frame is given by

$$E_v^f = \{e \in E : (v, f) \in R_1 \wedge (f, e) \in R_2\}. \tag{1}$$

Referring to Figure 1, the elements associated with the frame DEPARTING and the verb *leave* in particular could be represented as follows:

$$E_{leave}^{DEPARTING} = \{SOURCE, THEME\}.$$

To establish the empirical distribution of frame elements in a corpus, all observed elements need to be counted (cf. Definition 2.1.1), giving rise to a discrete probability distribution as in Definition 2.1.2. In order to obtain the observed frequencies and probabilities, an extraction pipeline has been developed using Python’s Natural Language Toolkit (NLTK) FrameNet corpus (version 1.7) (Bird et al., 2009); the output was subsequently processed in RStudio (RStudio Team, 2020; R version 4.4.2) and Python (Van Rossum & Drake, 2009; Python version 3.13.1). The set of extracted lexical units is limited to verbs, but it could be readily extended to include further parts of speech (see Busse, 2012, p. 748). As part of this procedure, data on  $|V| = 2,679$  verbs have been obtained, co-occurring with  $|E| = 822$  distinct frame elements and representing  $|F| = 641$  frames in total. For each verbal unit, I processed all attested examples, extracted the annotated frame elements, and aggregated frequency counts

for each tuple  $(v, f, e)$ , yielding 21,160 distinct combinations with a total element count of 193,349.

**Definition 2.1.1** (Observed frequency). Let  $N_v^f$  denote the total number of times verb  $v$  evokes frame  $f$  in the corpus. The observed frequency  $O_v^f(e)$  of element  $e$  given verb  $v$  and frame  $f$  is the count of corpus instances where element  $e$  appears when verb  $v$  evokes frame  $f$ :

$$O_v^f(e) = \sum_{i=1}^{N_v^f} [\text{element } e \text{ occurs in instance } i] \tag{2}$$

**Definition 2.1.2** (Probability mass function). The probability mass function  $P_v^f(e)$  returns the probability that element  $e$  occurs given verb  $v$  and frame  $f$  with  $\sum_{e \in E_v^f} P_v^f(e) = 1$ . The maximum likelihood estimate is given by

$$P_v^f(e) = \frac{O_v^f(e)}{\sum_{e \in E_v^f} O_v^f(e)}. \tag{3}$$

Elements that are theoretically possible but not observed in the corpus (i.e.,  $O_v^f(e) = 0$ ) are excluded from the support of the probability mass function.

### 2.2. Frame elements and Zipf’s law

For illustration, the frame element distribution of *eat* (INGESTION) is presented in Table 1. Note that  $z$  denotes the (Zipfian) frequency rank of an element, which is inversely proportional to its observed frequency. In each case, there are two high-probability elements which contrast with a longer tail of considerably rarer ones.

The skewed patterns of these two frame-evoking elements might reflect a more general distributional regularity. In fact, Liu (2011, pp. 211–212) has previously shown that the syntactic complements of English verbs obey a power law. As frame elements can be viewed as semantic complements, it can be hypothesised that frame elements follow a power law as well. By analogy with Baayen (2001, p. 20), one would need to show that the power model

$$O_v^f(e) = az^{-b} \tag{4}$$

reasonably fits the observed data. If the logarithm is taken over both sides, then the respective quantities should be linear in the log–log plane:

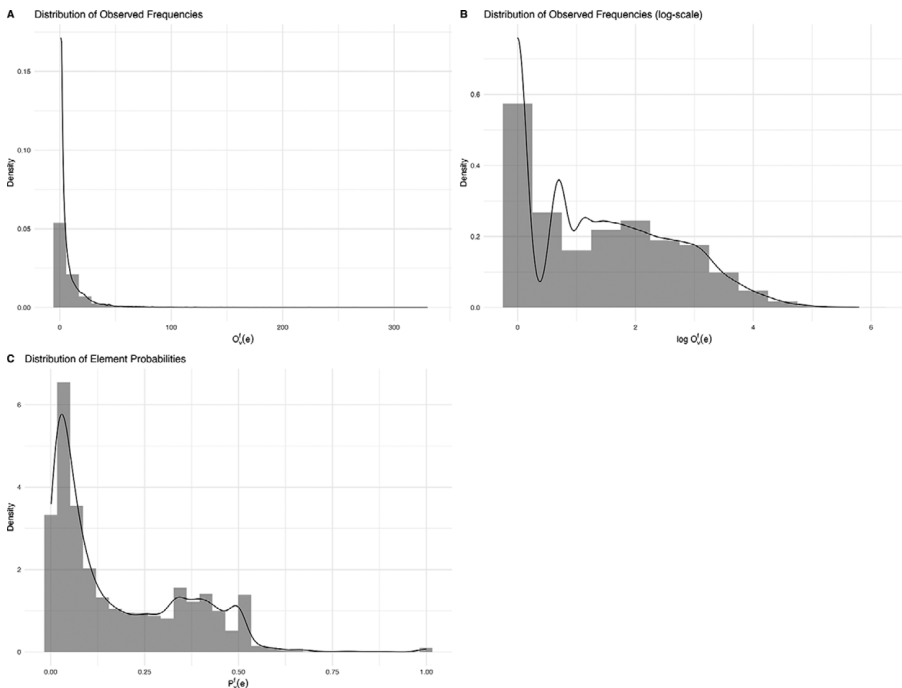
**Table 1.** Frame element distribution of the verb *eat* in the INGESTION frame with  $N = 35$

$e$	Coreness	$z$	$O(e)$	$P(e)$
INGESTOR	Core	1	24	0.41
INGESTIBLES	Core	2	23	0.40
PLACE	Non-core	3	5	0.09
MANNER	Non-core	4	3	0.05
SOURCE	Non-core	5	2	0.03
TIME	Non-core	6	1	0.02

$$\log O_v^f(e) = \log a + b \log z. \quad (5)$$

For instance, regressing  $\log z$  over  $\log O_v^f(e)$  returns an  $R^2$  of 0.89 for the distribution in Table 1. Repeating this procedure for the 3,875 frame element distributions  $E_v^f$  represented in the FrameNet corpus, least squares solutions could be computed for 3,219. For these cases, it can be confirmed that frame elements follow a Zipfian power law rather well (Mean  $R^2 = 0.82$ , SD = 0.11).<sup>8</sup>

General distributional characteristics of  $O_v^f(e)$  and  $P_v^f(e)$  are visualised in Figure 2. Panel A shows the distribution of observed frequencies on a linear scale, exhibiting extreme right-skew, with the vast majority of observations concentrated at very low frequencies. The density curve peaks sharply near zero and decays rapidly, with most frame elements occurring fewer than 50 times. The long tail extends to the most frequent frame element ( $O_v^f(e) = 330$ ), yet only 25% of elements have a frequency greater than 10. Panel B represents the same frequency distribution on a log-scale, revealing some additional structure. The log-frequencies show a more complex distribution with multiple local modes. This pattern suggests distinct ‘tiers’ of element frequencies, approximately binning them into rare, common and highly frequent elements.



**Figure 2.** Density plots of observed (log-)frequencies and sample proportions in the FrameNet data. A: Distribution of observed frequencies. B: Distribution of observed frequencies (log-scale). C: Distribution of element probabilities.

<sup>8</sup>Since the determination coefficient is positively correlated with  $N_v^f$  (Pearson's  $r = 0.48$ , 95% CI [0.45, 0.50],  $t(3217) = 30.91$ ,  $p < 0.001$ ), the relative fit is generally better if more frame elements are attested.

Unsurprisingly, the element probabilities in Panel C similarly demonstrate heavy concentration in the lower value range. The density is highest near zero, with most element having probabilities below 0.10. The distribution continues to show a deep decline followed by several smaller peaks around 0.30–0.50 and approximately 0.60. The long right tail extends to probability values near 1.0, though such cases are very rare.

All of these observations are consistent with pronounced power law behaviour. Regardless, these patterns pose challenges for further statistical inference and, thus, for capturing coreness. Specifically, the extreme sparsity and log-linear structure of frame element distributions call for measures that (a) handle heavily skewed discrete data and (b) respect the logarithmic scaling inherent in power law phenomena. In this regard, information theory provides an apt toolkit for the analysis of argument realisation.

### 2.3. Information-theoretic framework

Information-theoretic measures are regularly applied in a variety of linguistic sub-disciplines, including corpus linguistics (Gries, 2020), natural language processing (Manning & Schütze, 1999), syntactic typology (Levshina, 2019) and psycholinguistics (e.g., Frank et al., 2015; Linzen & Jaeger, 2016; Smith & Levy, 2013), among others. Recent advances include information-theoretic studies of valency (Say, 2025). Since argument realisation is regularly viewed in connection with predictability (e.g., Givón, 2017, p. 187 or Goldberg, 2001, p. 519), and information-theoretic metrics can capture the relative predictability of discrete categories while operating naturally in log-space, they constitute suitable measures for the analysis of frame element distributions. In short, information-theoretic measures represent ‘[d]istributional information for language’ (Slaats & Martin, 2025, p. 234), which is precisely what is needed for a usage-based characterisation of frame element coreness.

To introduce the basic concepts, recall the set  $E_v^f$ , which is a set of discrete categories, and the probability mass function  $P_v^f(e)$ , which returns the probability of observing an element  $e \in E_v^f$ . A fundamental measure in information theory is surprisal, which quantifies how unexpected, uninformative, or surprising a particular outcome is (Cover & Thomas, 1991, p. 13; Willems et al., 2016, p. 2507). The surprisal (or information content) of a singular frame element  $S_v^f(e)$  is defined as follows:

**Definition 2.3.1** (Surprisal of frame elements). If  $P_v^f(e)$  denotes the probability that a frame element  $e$  occurs with verb  $v$  and  $f$ , then the surprisal of said element is:

$$S_v^f(e) := -\log_2 P_v^f(e) \text{ with } \log_2 0 = 0. \quad (6)$$

This value, measured in bits if the logarithmic base is 2, increases as the probability  $P_v^f(e)$  decreases: Rare (less probable) elements are more informative when they occur and thus have higher surprisal values, as opposed to more common outcomes. As a natural consequence of the log-scaling, small differences in probability are strongly amplified. While the difference between the probabilities 0.03 and 0.01 appears minute (cf. MANNER vs. SOURCE in Table 1), their surprisal equivalents are  $-\log_2(0.0345) = 3.37$  and  $-\log_2(0.0172) = 4.06$ , respectively. Because frame elements follow a power law and their corresponding probabilities exhibit a strong right-skew, the majority of the probability mass involves low probability elements with

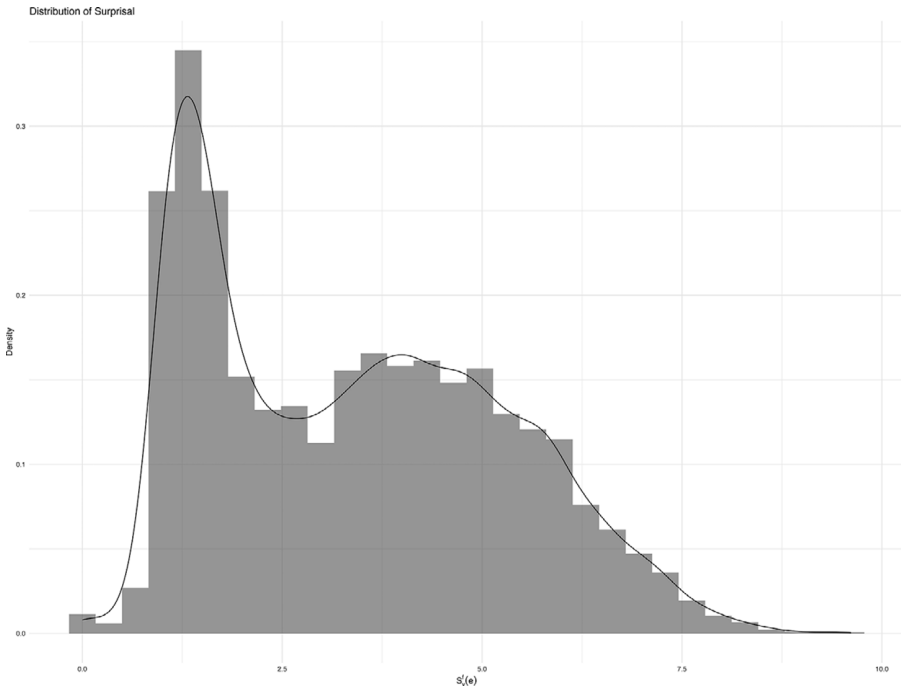


Figure 3. Density plot of surprisal in the FrameNet data.

small differences in values; surprisal effectively ‘uncompresses’ the probability distribution. This is particularly striking in Figure 3, which exhibits not only more symmetry but also clearer bimodality compared to the untransformed proportions in Figure 2C. Two clusters emerge: uninformative, high probability elements (left peak), and informative, moderate-to-low probability elements (right peak), potentially corresponding to core and non-core elements, respectively.

When the weighted sum is taken across all surprisal values in a frame element distribution, we obtain its Shannon entropy  $H(E_v^f)$  (cf. Definition 2.3.2), which, in the current application, measures the average information contained in the distribution of frame elements for a given verb  $v$  and frame  $f$ . This is because Shannon entropy equals the mathematical expectation  $\mathbb{E}$  of  $S_v^f(e)$ , written  $\mathbb{E}[S_v^f(e)]$ .<sup>9</sup> Since each element’s contribution to the overall entropy is weighted by its occurrence probability, high-probability elements will minimally increase uncertainty, whereas unlikely ones will contribute more substantially. Entropy reaches its maximum value when all outcomes are equally likely (Shannon, 1948), representing the greatest amount of uncertainty in outcomes.<sup>10</sup> Entropy is low if ‘there are few options, or

<sup>9</sup>The expectation of a discrete random variable  $X$  is defined as follows (cf. Ross, 2010, p. 36; Baayen, 2001, p. 37):

$$\mathbb{E}[X] = \sum_x xP(X = x). \quad (7)$$

<sup>10</sup>To be precise, its upper bound corresponds to the logarithm of the number of elements in the distribution, i.e.,  $\log|E_v^f|$ ; for the proof, see Cover and Thomas (1991), p. 27.

one of them has a much higher probability than the others' (Slaats & Martin, 2025, p. 234). In FrameNet, the entropy of frame element distributions exhibits some bimodality (cf. Figure C1A), indicating that there is a noteworthy set of frames with little uncertainty ( $H_v^f \approx 1$ ).<sup>11</sup> As a distribution-level feature, entropy characterises semantic valency patterns as a whole, revealing the degree of asymmetry in argument realisation preference.

**Definition 2.3.2** (Entropy of a frame element distribution). Suppose every element  $e \in E_v^f$  has probability  $P_v^f(e)$  and surprisal  $S_v^f(e)$ . The entropy of the corresponding frame element distribution, written  $H(E_v^f)$ , is:

$$H(E_v^f) = - \sum_{e \in E_v^f} P_v^f(e) \log_2 P_v^f(e). \quad (8)$$

A useful consequence of Definitions 2.3.1 and 2.3.2 is that the surprisal of each individual outcome  $e \in E_v^f$  can be compared to the average surprisal of the whole distribution. That is, one can calculate a deviation score that captures whether a frame element is more or less informative than expected in its distribution. This can be achieved in a way that respects the sign of the difference (cf. Definition 2.3.3) or instead emphasises the absolute difference by analogy with squared errors (cf. Definition 2.3.4). Some general statistical properties of (squared) entropy deviation are given in Appendix A. By analogy with the other measures, entropy deviation exhibits a bimodal distribution in the FrameNet data (Figure C1B), with a substantial portion of elements clustering slightly below 0.

**Definition 2.3.3** (Entropy deviation). Let  $S_v^f(e)$  denote the surprisal of a frame element and  $H(E_v^f)$  the entropy of its frame element distribution. Then, entropy deviation expresses how much more (or less) surprising a specific outcome  $e$  is upon occurrence compared to the expected level of surprise in the distribution:

$$\Delta_v^f(e) := S_v^f(e) - H(E_v^f). \quad (9)$$

**Definition 2.3.4** (Squared entropy deviation). Suppose  $\Delta_v^f(e)$  captures the entropy deviation of an element  $e$ . The squared entropy deviation captures how strongly this element deviates from  $H(E_v^f)$ :

$$[\Delta_v^f(e)]^2 := [S_v^f(e) - H(E_v^f)]^2. \quad (10)$$

The total spread of surprisal values around their expected value is their variance and is known as varentropy; in other words, it 'measures the variability in the information content' (Di Crescenzo & Paolillo, 2021, p. 682) of a random variable:

**Definition 2.3.5** (Varentropy). The total variance of surprisal values in a frame element distribution is the varentropy  $\mathbb{V}[S_v^f(e)]$ .

It can be easily shown that varentropy is the expected (i.e., average) squared deviation of surprisal values from the distributional average (cf. Proposition A.2). When comparing their corresponding histograms (cf. Figure C1C and C1D), frame

<sup>11</sup>The prevalence of low-entropy frames may be a consequence of data sparsity.

elements typically cluster quite closely around the expected surprisal (entropy), suggesting little deviation overall.

#### 2.4. *Word vectors*

Distributional semantic models represent word meaning on the basis of usage patterns in large corpora and have proven highly informative in the synchronic and diachronic study of English lexis, morphology and syntax (e.g., Hamilton et al., 2016; Marelli & Baroni, 2015; Perek, 2016; Shafaei-Bajestan et al., 2022). These approaches are fundamentally usage-based in that semantic content is approximated via statistical word co-occurrence patterns in representative collections of authentic linguistic data (Boleda, 2020, pp. 214–215; Günther et al., 2019).<sup>12</sup>

In the present study, distributional semantic representations are used to characterise the semantic profiles of FrameNet frame elements. While information-theoretic measures capture how frequently elements occur in particular frames, word embeddings may capture what semantic content typically fills a given frame element. Rather than modelling concrete phrasal or clausal realisations directly, the analysis abstracts away from surface structure and focuses on the aggregate semantic properties of frame element fillers. To this end, pre-trained word embeddings from the fastText library (Bojanowski et al., 2017; Mikolov et al., 2018; Mouselimis, 2024) were extracted for all attested frame element fillers in the FrameNet database. These embeddings incorporate subword information and are therefore well suited to modelling morphologically complex and low-frequency items.

At the level of individual corpus instances, frame elements realised by a single word are represented by that word's embedding, while multi-word realisations are represented by the unweighted average of the embeddings of their constituent words. This procedure reflects the assumption that words jointly realising a frame element contribute equally to its semantic interpretation, and it avoids introducing additional parameters for which no independent weighting criteria are available. To obtain distributional representations at the level required for the present analysis, these token-level representations are further aggregated across all attestations of a given verb–frame–element combination. The resulting frame element embeddings thus capture the typical distributional semantic profile of a frame element as it is realised with a particular verb and frame. A fully formal specification of the embedding construction and aggregation procedure is provided in [Appendix B](#).

#### 2.5. *Machine learning methods*

The statistical analysis in this study uses ensembles of boosted decision trees with random effects, implemented via the Gaussian Process Boosting (GPBoost) framework (Sigrist, 2022). This approach combines the strengths of tree-based machine learning (such as decision trees and random forests, which are already familiar in corpus linguistics; e.g., Tagliamonte & Baayen, 2012) with mixed-effects modelling. In essence, GPBoost enables the modelling of complex, non-linear relationships

<sup>12</sup>This assumption is commonly referred to as the Distributional Hypothesis; see Lenci (2008) and Firth (1968) for detailed discussion.

between predictors and outcomes while also accounting for hierarchical structure in the data (Sigrist, 2022, p. 1).

For each coreness distinction, I fit a mixed-effects model with both fixed and random effects and compare it to a null model with random effects only. Information-theoretic measures and word vectors constitute the fixed effects, and ‘verb’ as well as ‘frame’ the grouping factors. Formally, the model predicts the response vector  $\mathbf{y}$  (e.g., whether an element is core or non-core) following Equation (11).

$$\mathbf{y} = F(\mathbf{X}) + \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon} \quad \text{with } \mathbf{b} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}), \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_n) \quad (11)$$

$F(\mathbf{X})$  is the set of boosted trees used to approximate the relationship between the independent variables  $\mathbf{X}$  and the target variable,  $\mathbf{Z}\mathbf{b}$  represents random effects, and  $\boldsymbol{\epsilon}$  is the model error. The grouping factor is assumed to follow a multivariate normal distribution with mean vector  $\boldsymbol{\mu} = \mathbf{0}$  and covariance matrix  $\boldsymbol{\Sigma}$ , and the residuals should be independent and identically distributed normal variables.

Model fitting in GPBoost proceeds via gradient boosting, where each decision tree is sequentially added to minimise prediction error by following the direction of steepest descent in the loss function (gradient descent; see Hastie et al., 2017, pp. 358–360, for details). To ensure robust predictive performance, hyperparameters (such as tree depth and learning rate) are optimised using randomised grid search with binary log-loss (cross-entropy) as the minimisation criterion. All models are trained and validated in Python’s scikit-learn environment (Pedregosa et al., 2011), with the data split into 75% training and 25% test sets to assess overfitting.

Predictive performance is evaluated using standard classification metrics, namely, accuracy, AUC, precision, recall and the F1-score; their calculation is summarised in Table 2. In addition, the Matthews Correlation Coefficient (MCC, see Equation 12)

**Table 2.** Summary of common classification metrics (TP = true positive, TN = true negative, FP = false positive, FN = false negative) based on Chicco and Jurman (2023, pp. 2–3), Alpaydin (2022, pp. 632–636) and James et al. (2021, pp. 148–152)

Metric	Formula
Positive predictive value/precision	$\frac{TP}{TP + FP}$
TP rate/recall/sensitivity	$\frac{TP}{TP + FN}$
TN rate/specificity	$\frac{TN}{TN + FP}$
Negative predictive value	$\frac{TN}{TN}$
FP rate	$\frac{FP}{FP + TN}$
FN rate	$\frac{FN}{FN + TN}$
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$
F1	$2 \times \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$
ROC	Recall vs. False Positive Rate
AUC	Area under the ROC curve

will be reported by analogy with Yoffe et al. (2025, p. 10). The MCC is the confusion matrix analogue of Pearson's  $r$ , ranging from  $-1$  (all labels misclassified) to  $+1$  (all labels correctly classified), with  $0$  representing the random baseline (Chicco & Jurman, 2020, p. 5). It can only be high if both positive cases are consistently identified correctly (i.e., high precision and recall) and negative ones are as well. As such, for each of the three coreness contrasts, it will provide a joint metric of how well both outcomes can be predicted.

$$\text{MCC} = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP) \cdot (TP + FN) \cdot (TN + FP) \cdot (TN + FN)}} \quad (12)$$

### 3. Results

The information-theoretic measures introduced in Section 2.3 can be harnessed to capture fine-grained quantitative characteristics of the right-skewed frame element distributions. Table 3 revisits the INGESTION frame and provides information-theoretic metrics for each element. It is worth noting that some elements appear to be less surprising than expected upon occurrence ( $\Delta_v^f < 0$ ), while some are more surprising ( $\Delta_v^f > 0$ ). At the same time, two elements, which are in fact core elements of this frame, are very close to the expected surprisal of elements in this frame ( $[\Delta_v^f]^2 \approx 0$ ), while the non-core elements are substantially further away ( $[\Delta_v^f]^2 \gg 0$ ). This pattern is remarkably consistent across a variety of verb-frame combination (cf. Figure 4), albeit not categorical; some putative core elements (e.g., DELIVERY\_DEVICE in Panel E) pattern more closely to non-core elements.

#### 3.1. Descriptive overview

The frame element types are distributed unevenly (see Table 4), with core elements constituting half of all realised elements, peripheral ones a third and extra-thematic ones approximately one-sixth. Core-unexpressed elements are rare at less than 1% of occurrences. Table 5 presents descriptive statistics on the observed usage data. Particularly striking is the asymmetry and extreme tailedness in the observed counts: Typically, the element distributions derived from the FrameNet data involve low-frequency and low-probability elements while still allowing for substantial outliers. This pattern aptly reflects the data reported in Table 3 earlier. It also applies to the

**Table 3.** Probability distribution of frame elements conditioned on the verb *eat* and the frame INGESTION with a Shannon entropy  $H(E_{eat}^{\text{INGESTION}})$  of 1.85 bits and a varentropy of 1.38

Element	$O(e)$	$P(e)$	$S(e)$	$\Delta(e)$	$[\Delta(e)]^2$
INGESTOR <sub>core</sub>	24	0.41	1.27	-0.58	0.33
INGESTIBLES <sub>core</sub>	23	0.40	1.33	-0.52	0.27
PLACE <sub>non-core</sub>	5	0.09	3.54	1.69	2.84
MANNER <sub>non-core</sub>	3	0.05	4.27	2.42	5.87
SOURCE <sub>non-core</sub>	2	0.03	4.86	3.01	9.05
TIME <sub>non-core</sub>	1	0.02	5.86	4.01	16.10

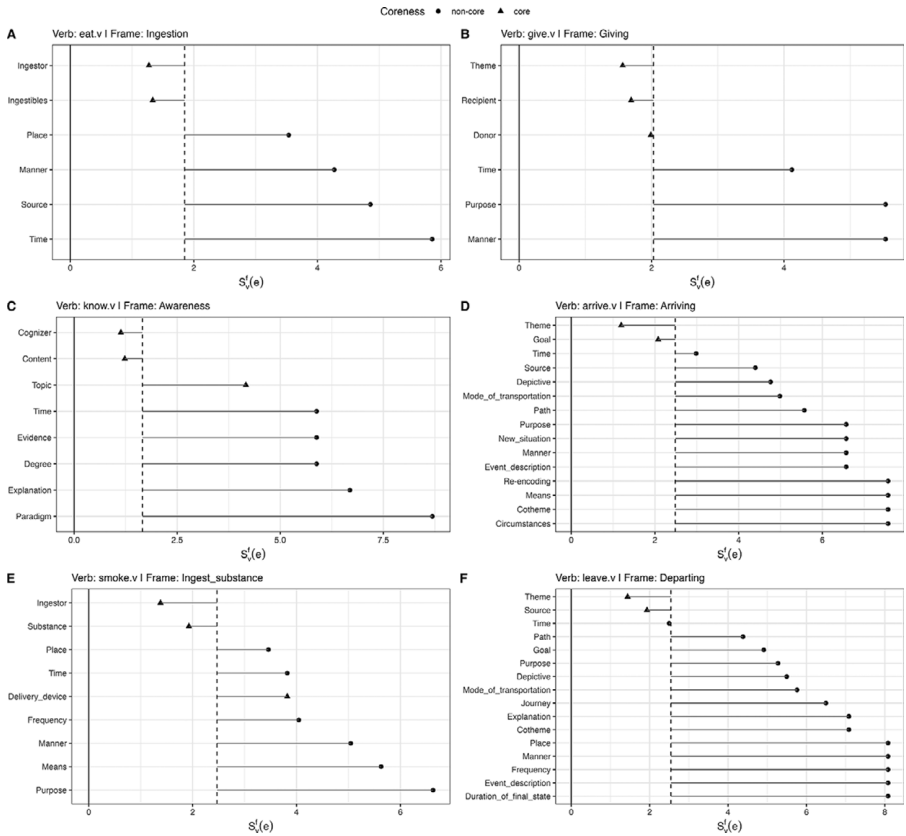


Figure 4. Surprisal  $S_v^f(e)$ , Shannon entropy (dashed line) and entropy deviation (solid horizontal lines) of frame element types conditioned on verbs in different frames.

Table 4. Distribution of element types for all verbs in the FrameNet database

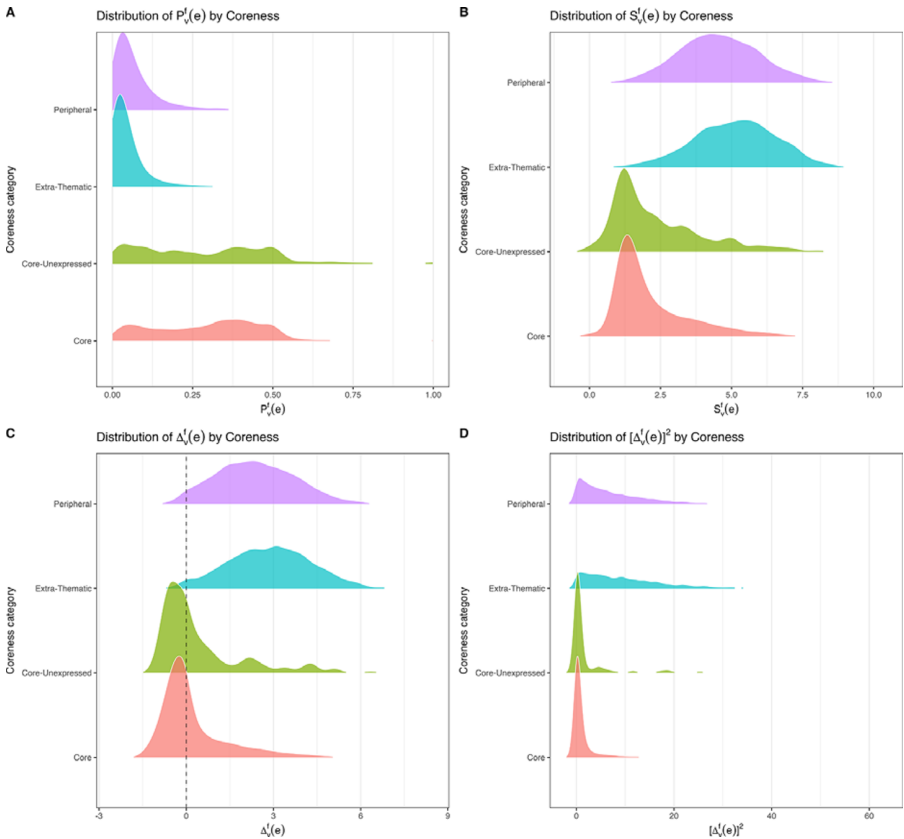
Coreness	Count	Pct (%)
Core	10694	50.54
Core-unexpressed	188	0.89
Extra-thematic	3098	14.64
Peripheral	7180	33.93
<b>Total</b>	<b>21160</b>	<b>100</b>

squared entropy deviation, albeit in a strongly attenuated form, as the bulk of elements tend to be rather close to the expected surprisal of their distribution.

Figure 5 illustrates the conditional distributions of element probabilities and information-theoretic measures by element coreness. Given that counts are sample-dependent, the estimated proportions are provided in Figure 5A. These data show that the vast majority of non-core elements show low probabilities of occurrence, with only one out 10,278 exceeding the 0.5 probability mark. By contrast, core elements are considerably more spread out, taking on values across the entire

**Table 5.** Summary statistics on the predictor variables ( $n = 21,160$ )

Statistic	$O_v^f(e)$	$P_v^f(e)$	$S_v^f(e)$	$H(E_v^f)$	$\Delta_v^f(e)$	$[\Delta_v^f(e)]^2$	$\mathbb{V}[S_v^f]$
Mean	9.49	0.18	3.41	2.05	1.36	4.88	1.20
SD	16.43	0.18	1.90	0.54	1.74	7.11	0.68
Median	4.00	0.10	3.32	2.06	1.09	1.40	1.20
MAD	4.45	0.12	2.49	0.53	2.00	2.06	0.72
Min	1.00	0.00	0.00	0.00	-1.88	0.00	0.00
Max	377.00	1.00	9.63	3.65	7.84	61.43	3.49
Skew	5.56	1.04	0.38	-0.23	0.54	2.14	0.19
Kurtosis	56.13	0.71	-0.89	0.21	-0.64	5.66	-0.40



**Figure 5.** Distribution of information-theoretic measures across frame element types (density plots). A: Distribution of  $P_v^f(e)$  by coreness. B: Distribution of  $S_v^f(e)$  by coreness. C: Distribution of  $\Delta_v^f(e)$  by coreness. D: Distribution of  $[\Delta_v^f(e)]^2$  by coreness.

[0, 1] interval, with 75% lying below 0.42. Accordingly, core and core-unexpressed elements concentrate around lower surprisal values (Panel 5B), which correspond to probabilities of occurrence close to 1, but display negative skew with long tails. Non-core elements show more symmetric distributions with notably higher surprisal

values; for reference, a surprisal of 5 corresponds to a probability of  $2^{-5} = 0.03$  and a value of 7.5 to  $2^{-7.5} = 0.006$ .

When examining the entropy deviations (Panel C), core elements cluster closely around the average surprisal of their distributions, with the majority having deviation scores less than 0, reflecting the intuition that core elements are less surprising than expected upon occurrence. Extra-thematic and peripheral elements consistently show positive entropy deviations, carrying more information than average. The squared deviation (Panel D) highlights the magnitude of these deviations regardless of direction. On the whole, the squared deviation scores are right-skewed, with a small subset of frame elements exhibiting exceptionally large departures from expected surprisal values. Non-parametric Dunn tests reveal no significant differences between core and core-unexpressed elements across any information-theoretic measure, while all other pairwise comparisons between coreness types are statistically significant (cf. Table 6).

Prior to model fitting, it is worth inspecting the pairwise feature correlations in Table 7 to rule out mutual redundancies. In fact, the determinant of the matrix is 0, rendering it singular and non-invertible. Due to perfect collinearity between

**Table 6.** Dunn test results with Bonferroni-adjusted  $p$ -values ( $p_{\text{adjusted}}$ ) for pairwise comparisons between frame element types based on three entropy-derived metrics

Metric	Group 1	Group 2	$Z$	$p_{\text{adjusted}}$
$S_v^f(e)$	Core	Core-unexpressed	-1.43	0.920
	Core	Extra-thematic	-76.01	< 0.001
	Core-unexpressed	Extra-thematic	-19.25	< 0.001
	Core	Peripheral	-85.00	< 0.001
	Core-unexpressed	Peripheral	-16.13	< 0.001
	Extra-thematic	Peripheral	11.81	< 0.001
$\Delta_v^f(e)$	Core	Core-unexpressed	-1.53	0.759
	Core	Extra-thematic	-74.89	< 0.001
	Core-unexpressed	Extra-thematic	-18.85	< 0.001
	Core	Peripheral	-86.04	< 0.001
	Core-unexpressed	Peripheral	-16.25	< 0.001
	Extra-thematic	Peripheral	10.01	< 0.001
$[\Delta_v^f(e)]^2$	Core	Core-unexpressed	-1.54	0.744
	Core	Extra-thematic	-69.49	< 0.001
	Core-unexpressed	Extra-thematic	-17.37	< 0.001
	Core	Peripheral	-77.04	< 0.001
	Core-unexpressed	Peripheral	-14.38	< 0.001
	Extra-thematic	Peripheral	11.27	< 0.001

**Table 7.** Feature correlation matrix based on Spearman's  $\rho$

	$O_v^f(e)$	$P_v^f(e)$	$S_v^f(e)$	$H(E_v^f)$	$\Delta_v^f(e)$	$[\Delta_v^f(e)]^2$	$\mathbb{V}[E_v^f]$
$O_v^f(e)$							
$P_v^f(e)$	0.623						
$S_v^f(e)$	-0.623	-1.000					
$H(E_v^f)$	0.019	-0.439	0.439				
$\Delta_v^f(e)$	-0.723	-0.937	0.937	0.146			
$[\Delta_v^f(e)]^2$	-0.560	-0.883	0.883	0.277	0.842		
$\mathbb{V}[E_v^f]$	0.041	-0.509	0.509	0.597	0.337	0.526	

probability and surprisal, and given the information-theoretic approach pursued here, surprisal is retained as the more theoretically motivated measure of element predictability. Moreover, raw observed counts were excluded due to their sample dependence, as absolute frequencies may reflect corpus composition rather than inherent usage patterns. The analysis instead focuses on distributional properties that capture relative predictability across different verb-frame combinations.

### 3.2. Model evaluation

Tables 8 and 9 compare the overall predictive performance of the mixed-effects gradient boosting models with and without usage data. The corresponding metrics are visualised in Figure 6.<sup>13</sup> Usage data provide substantial improvements for the core versus non-core distinction, with MCC increasing from 0.542 to 0.817 and accuracy from 0.527 to 0.907. As for the peripheral versus extra-thematic contrast, the MCC only improves marginally (MCC: 0.533 → 0.556). The core versus core-unexpressed distinction shows high accuracy in both models (> 0.98), though this largely reflects class imbalance given the extreme rarity of core-unexpressed elements (0.89% of cases). Quite strikingly, the MCC for this contrast decreases from 0.813 to 0.538 with the addition of usage data, indicating that the null model's apparent success was driven by predicting the majority class rather than genuine discriminative power.

Class-specific metrics (Tables 10 and 11, Figure 7) reveal distinct patterns across contrasts. For core versus non-core, the full model achieves balanced performance (F1 ≈ 0.919 for both classes), while the null model defaults to predicting 'core' for nearly all instances (non-core recall = 0.040). This demonstrates that distributional information is essential for this distinction.

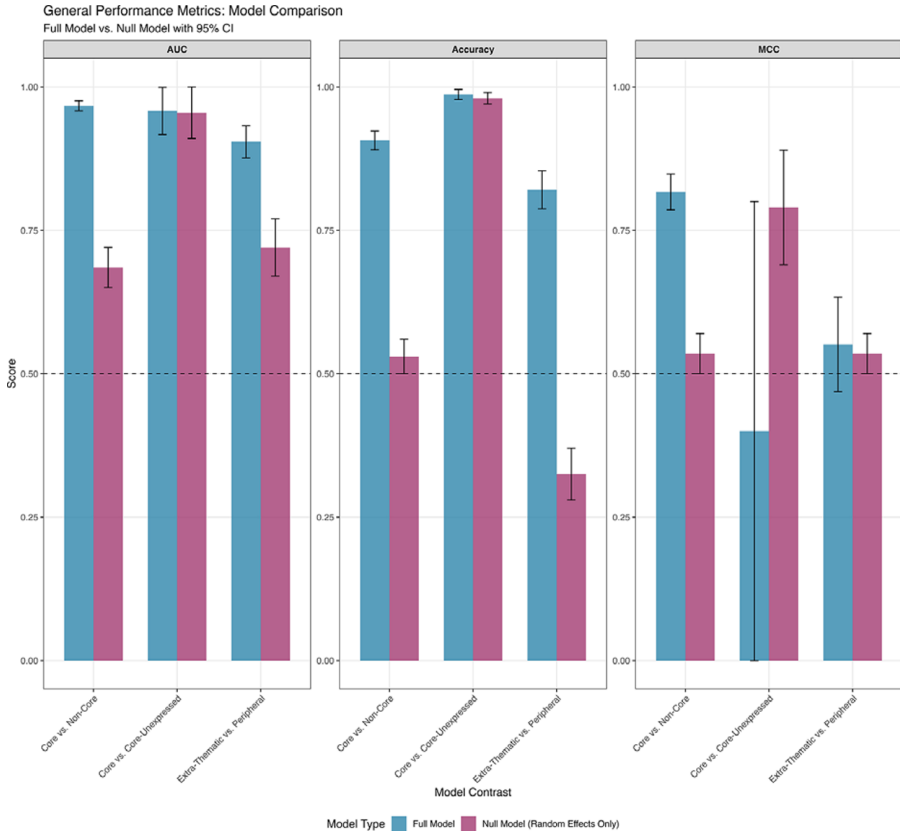
**Table 8.** General performance metrics for the full model with fixed and random effects (95% confidence intervals); MCC normalised to [0,1] scale

Metric	Core vs. Non-core	Core vs. Core-unexpressed	Peripheral vs. Extra-thematic
AUC	0.967 [0.958, 0.976]	0.977 [0.917, 0.999]	0.905 [0.877, 0.932]
Accuracy	0.907 [0.891, 0.923]	0.988 [0.978, 0.996]	0.821 [0.787, 0.854]
Precision	0.910 [0.895, 0.925]	0.987 [0.972, 0.996]	0.837 [0.805, 0.866]
Recall	0.907 [0.891, 0.923]	0.988 [0.978, 0.996]	0.821 [0.787, 0.854]
F1	0.907 [0.891, 0.923]	0.985 [0.972, 0.996]	0.799 [0.753, 0.839]
MCC	0.817 [0.786, 0.848]	0.538 [0.000, 0.800]	0.556 [0.469, 0.634]

**Table 9.** General performance metrics for the random-effects-only model (95% confidence intervals); MCC normalised to [0,1] scale

Metric	Core vs. Non-core	Core vs. Core-unexpressed	Peripheral vs. Extra-thematic
AUC	0.681 [0.651, 0.715]	0.973 [0.909, 0.997]	0.721 [0.667, 0.767]
Accuracy	0.527 [0.501, 0.559]	0.982 [0.971, 0.991]	0.319 [0.278, 0.372]
Precision	0.602 [0.258, 0.737]	0.988 [0.979, 0.995]	0.676 [0.083, 0.802]
Recall	0.527 [0.501, 0.559]	0.982 [0.971, 0.991]	0.319 [0.278, 0.372]
F1	0.387 [0.341, 0.432]	0.984 [0.974, 0.992]	0.176 [0.126, 0.258]
MCC	0.542 [0.500, 0.574]	0.813 [0.691, 0.892]	0.533 [0.500, 0.571]

<sup>13</sup>Note that the MCC has been rescaled to the interval [0,1] for interpretability.



**Figure 6.** Summary of general performance metrics (dashed line represents expected values if the model was guessing randomly).

The core versus core-unexpressed contrast presents a different challenge. While core elements are predicted near-perfectly ( $F1 = 0.994$ ), core-unexpressed elements exhibit extreme variance with confidence intervals spanning  $[0.000, 1.000]$  for both precision and recall. This instability persists even with usage data, suggesting insufficient training instances for stable pattern learning.<sup>14</sup> Counterintuitively, the full model's MCC (0.538) is substantially lower than the null model's (0.813), though both exhibit wide confidence intervals. This pattern suggests that while the null model achieves moderate balanced performance by relying on structural patterns alone, the addition of usage features destabilises predictions for the rare core-unexpressed class without providing sufficient discriminative information.

For peripheral versus extra-thematic, the full model shows asymmetric performance: Peripheral elements are identified accurately ( $F1 = 0.884$ , recall = 0.980), while extra-thematic elements show high precision (0.910) but poor recall (0.453). The model correctly identifies extra-thematic elements when it predicts this label but

<sup>14</sup>In an attempt to remedy this imbalance, stratified sampling was used to ensure core-unexpressed elements were always included in the training data despite their rarity.

**Table 10.** Class-specific performance metrics for the Full Model (95% confidence intervals)

Contrast	Class	Metric	Mean	95% CI
Core vs. non-core	Core	Precision	0.945	[0.924, 0.965]
		Recall	0.894	[0.868, 0.918]
		F1	0.919	[0.904, 0.933]
	Non-core	Precision	0.894	[0.868, 0.918]
		Recall	0.945	[0.918, 0.967]
		F1	0.919	[0.900, 0.933]
Core vs. core-unexpressed	Core	Precision	0.989	[0.979, 0.997]
		Recall	0.999	[0.994, 1.000]
		F1	0.994	[0.989, 0.998]
	Core-unexpressed	Precision	0.853	[0.000, 1.000]
		Recall	0.368	[0.000, 0.727]
		F1	0.488	[0.000, 0.800]
Extra-thematic vs. peripheral	Extra-thematic	Precision	0.910	[0.837, 0.974]
		Recall	0.453	[0.344, 0.553]
		F1	0.603	[0.505, 0.686]
	Peripheral	Precision	0.805	[0.769, 0.843]
		Recall	0.980	[0.961, 0.996]
		F1	0.884	[0.864, 0.907]

**Table 11.** Class-specific performance metrics for the random-effects-only model (95% confidence intervals)

Contrast	Class	Metric	Mean	95% CI
Core vs. non-core	Core	Precision	0.521	[0.496, 0.551]
		Recall	0.989	[0.977, 1.000]
		F1	0.682	[0.659, 0.707]
	Non-core	Precision	0.687	[0.000, 0.960]
		Recall	0.040	[0.000, 0.069]
		F1	0.075	[0.000, 0.127]
Core vs. core-unexpressed	Core	Precision	0.997	[0.991, 1.000]
		Recall	0.984	[0.974, 0.993]
		F1	0.991	[0.985, 0.996]
	Core-unexpressed	Precision	0.314	[0.250, 0.711]
		Recall	0.837	[0.556, 1.000]
		F1	0.607	[0.354, 0.778]
Extra-thematic vs. peripheral	Extra-thematic	Precision	0.308	[0.266, 0.348]
		Recall	0.997	[0.981, 1.000]
		F1	0.470	[0.421, 0.516]
	Peripheral	Precision	0.836	[0.000, 1.000]
		Recall	0.025	[0.000, 0.082]
		F1	0.048	[0.000, 0.150]

misclassifies over half as peripheral. The null model defaults to predicting ‘extra-thematic’ for most instances (recall = 0.997), failing to identify peripheral elements (recall = 0.025). Despite similar MCC values between models (0.533 vs. 0.556), the full model achieves substantially higher accuracy (0.821 vs. 0.319), indicating that usage data enable meaningful classification where the null model cannot distinguish the categories.

Finally, the covariance parameters for the three models differ only in terms of frames, as they were consistently zero for the verbs. The core/non-core model showed the smallest contribution of frames as a random effect (0.13), closely followed by the peripheral/extra-thematic one (0.17). The outcomes of the core/core-unexpressed

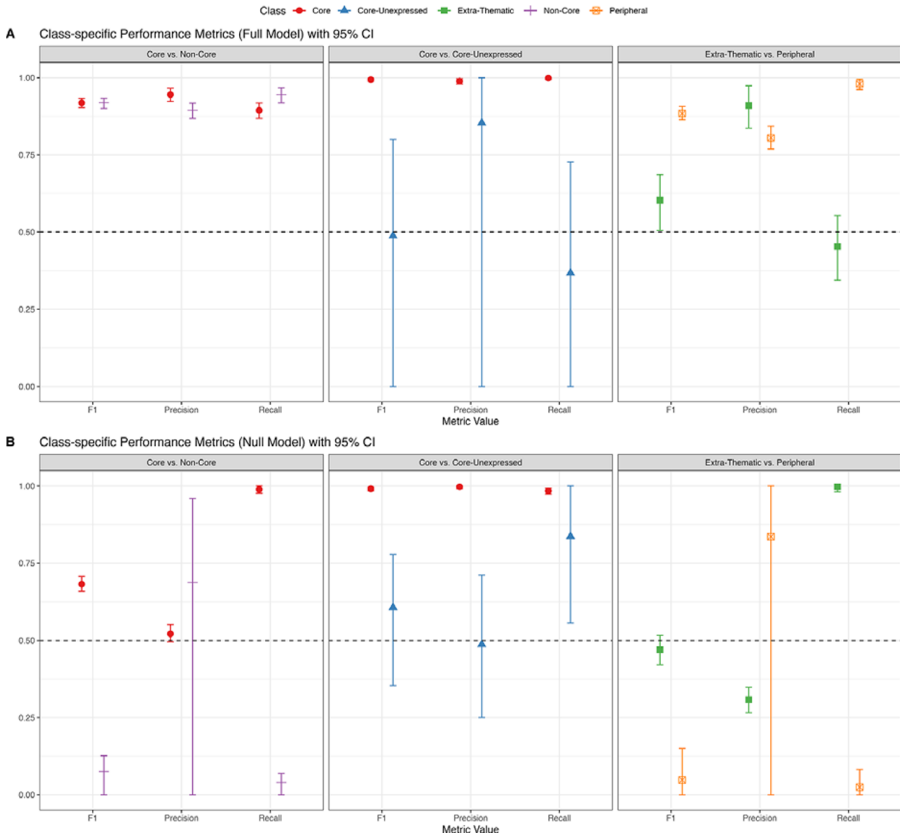


Figure 7. Summary of class-specific test performance with 95% CIs.

distinction, by contrast, were heavily affected by the frame under consideration (2.42), indicating that frame identity explains substantial between-frame variation in core-unexpressed classification.

### 3.3. Feature importance and partial dependence

The relative contributions of individual predictors are visualised in Figures 8 and 9. These plots include only predictors whose bootstrapped 95% confidence intervals exclude zero after 250 resampling iterations, indicating reliable non-zero contributions to model performance. While considerable overlap exists between confidence intervals across features, several patterns emerge. For the core versus non-core distinction (Figure 8), entropy deviation  $\Delta_v^f(e)$  and word vector dimension V163 emerge as the strongest predictors. Intermediate contributors include surprisal  $S_v^f(e)$ , V80 and V95. The peripheral versus extra-thematic distinction (Figure 9) is dominated by word vector dimensions V163, surprisal, and V134. No feature importance plot is presented for the core versus core-unexpressed contrast, as all predictors displayed confidence intervals spanning zero or including negative values, reflecting the instability discussed above.

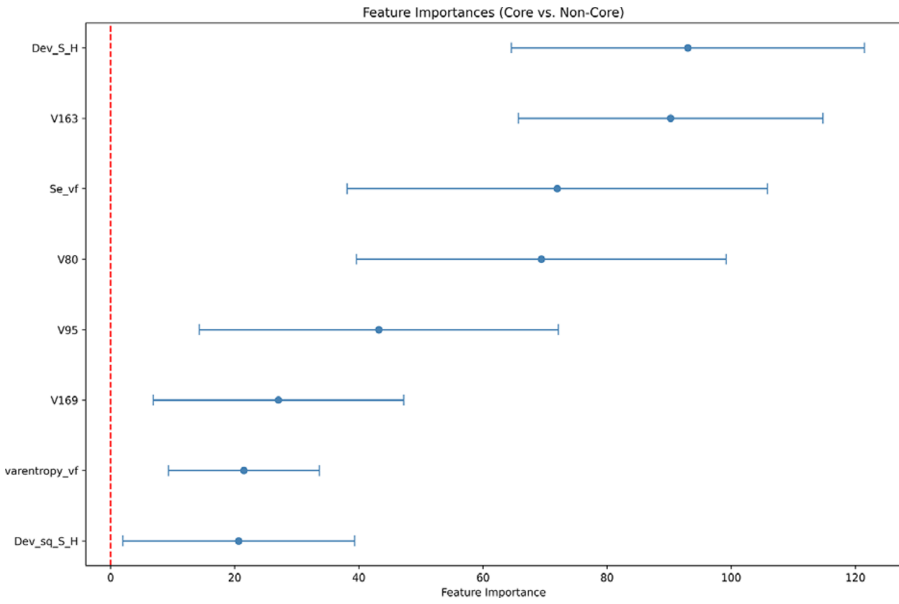


Figure 8. Feature importance scores for the core versus non-core distinction with bootstrapped 95% CIs ('H\_vf' =  $H(E_v^f)$ , 'Se\_vf' =  $S_v^f(e)$ , 'Dev\_S\_H' =  $\Delta_v^f(e)$ , 'Dev\_sq\_S\_H' =  $[\Delta_v^f(e)]^2$ ).

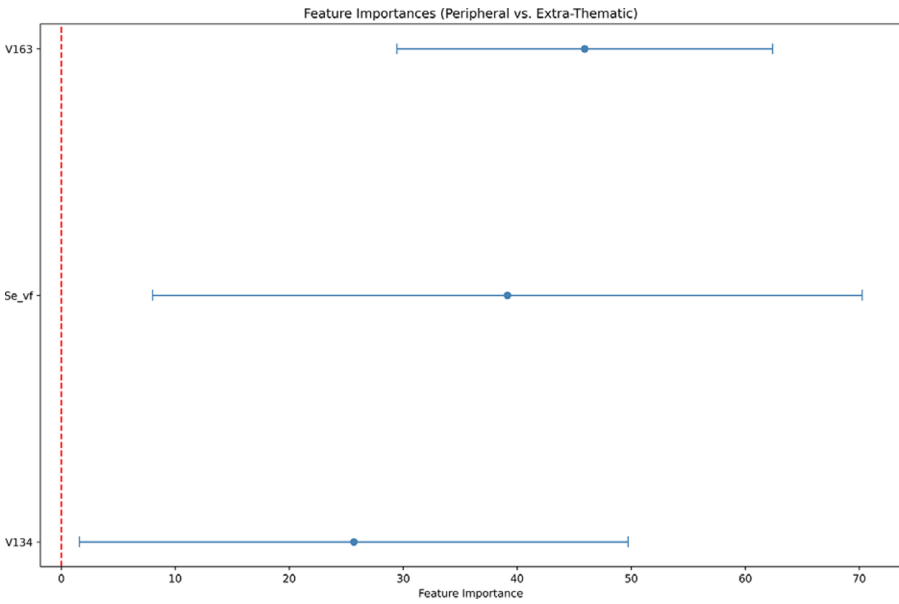
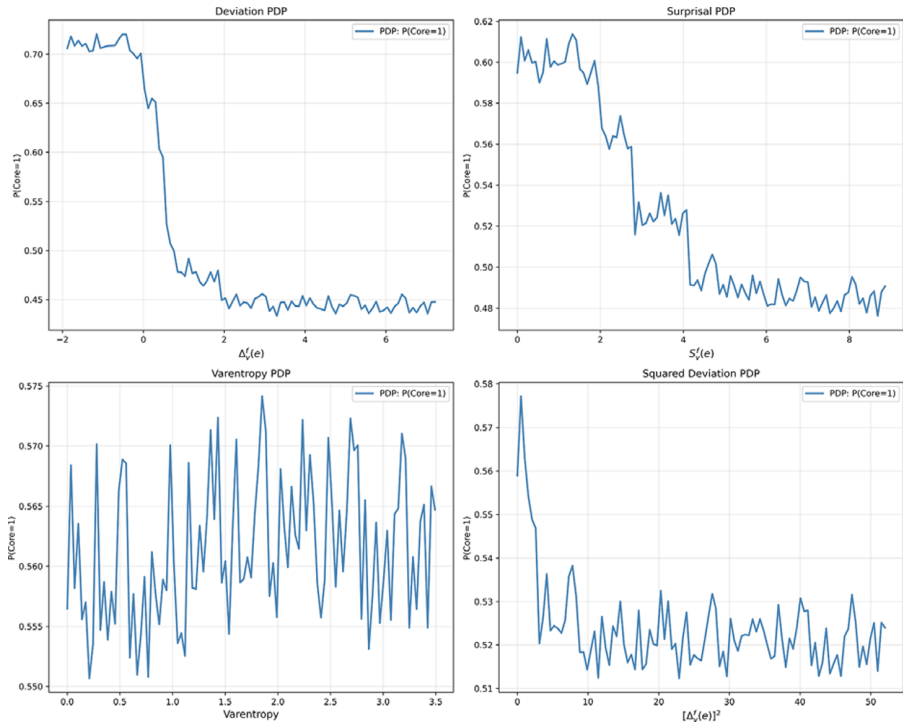


Figure 9. Feature importance scores for the extra-thematic versus peripheral distinction with bootstrapped 95% CIs.



**Figure 10.** Partial dependence of the core versus non-core model (information-theoretic measures).

Partial dependence plots (Figures 10–12) reveal the functional form of these relationships. For core versus non-core,  $\Delta_v^f(e)$  shows a robust positive effect: Non-core status becomes increasingly likely as an element’s surprisal exceeds the distribution’s entropy (i.e.,  $\Delta_v^f(e) > 0$ ). At  $\Delta_v^f(e) = 2$ , the predicted probability of ‘non-core’ is approximately 0.25 times higher than the model’s average prediction. Surprisal  $S_v^f(e)$  shows a weaker effect, with probability increases becoming relevant only for values  $\geq 3$ ; the effect of varentropy is markedly non-linear. Positive loadings of the word vector dimensions in Figure 11 appear to be associated with core status. Remarkably, higher V163 scores increase the probability that a frame element is a core element, and lower scores are linked to peripheral non-core elements. Surprisal appears to be marginally lower for peripheral elements compared to extra-thematic ones.

## 4. Discussion

### 4.1. Summary and limitations

The present study examined whether FrameNet’s coreness distinctions are grounded in actual language use by training machine learning models on information-theoretic measures and word embeddings to predict three coreness contrasts: (i) core versus non-core, (ii) core versus core-unexpressed and (iii) peripheral versus extra-thematic. The choice of information-theoretic predictors was motivated by the

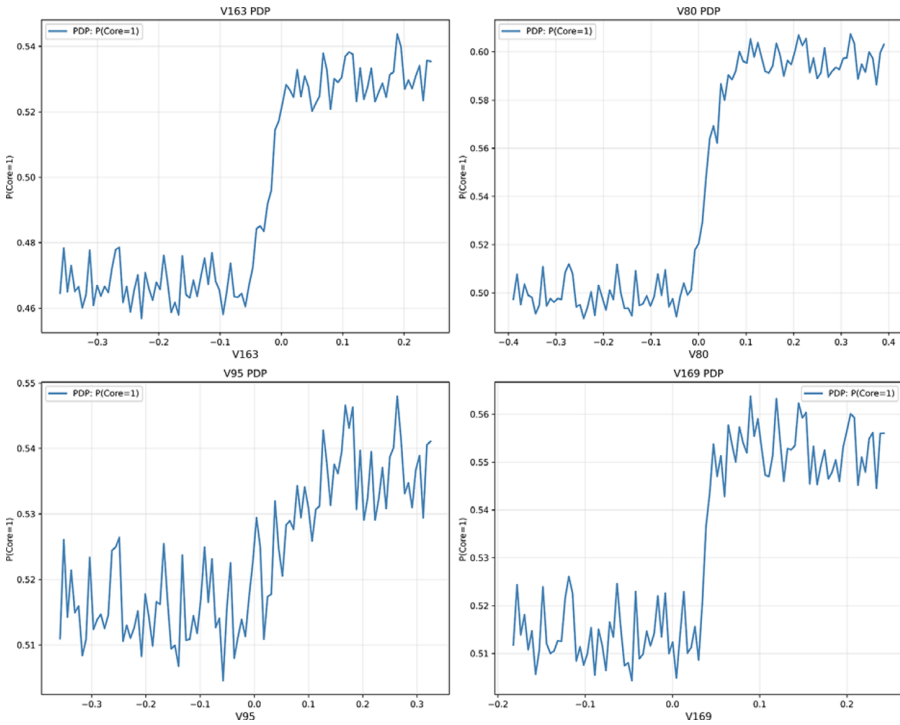
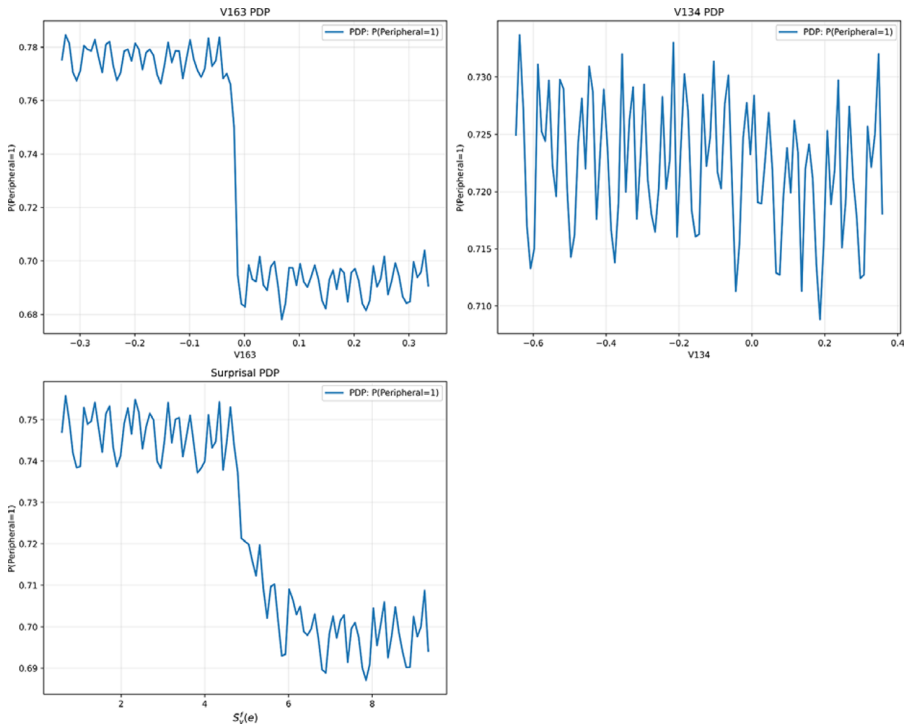


Figure 11. Partial dependence of the core versus non-core model (word vectors).

finding that frame elements follow Zipf’s law, making entropy-based measures particularly well-suited to capturing their predictability patterns. Word embeddings complement this approach by capturing the semantic profiles of frame element fillers grounded in authentic language use. The results provide clear but differentiated support for these distinctions:

- 1) The core versus non-core contrast showed strong usage-based grounding. The full model achieved near-perfect classification, with entropy deviation  $\Delta_v^f(e)$  emerging as the strongest predictor. Crucially, the null model containing only lexical and frame information failed catastrophically, demonstrating the importance of usage data. Partial dependence analysis additionally reveals important quantitative characteristics of core elements:
  - Core elements display  $\Delta_v^f(e) \leq 0$ . Their information content is systematically below the average (entropy) of the frame element distribution.
  - Core elements minimise  $S_v^f(e)$ . They are highly likely to occur and thus do not convey much new information upon occurrence.
  - Core elements minimise  $[\Delta_v^f(e)]^2$ . They cluster relatively closely around the entropy of the frame element distribution.
  - Core elements maximise word vectors V163, V80, V95 and V169. Thus, they exhibit characteristic distributional semantic profiles.
- 2) The peripheral versus extra-thematic distinction showed moderate usage-based support. The classifier relied on semantic features (word vectors V163 and V134) as well as on surprisal, which did not raise the general performance



**Figure 12.** Partial dependence of the extra-thematic versus peripheral model.

metrics above the random baseline. They did, however, improve the class-specific performance, especially the identification of peripheral elements. In particular, these display the following features:

- Peripheral elements minimise  $S_v^f(e)$ . Their information content is on the lower end, yet it is higher than that of core elements.
  - Peripheral elements minimise V163, and they display a non-linear association with V134.
- 3) The core versus core-unexpressed contrast proved empirically problematic. Despite high overall accuracy, this reflected class imbalance rather than genuine discriminative success. Confidence intervals for core-unexpressed precision and recall spanned [0.000, 1.000], indicating extreme instability across bootstrap samples. It is, therefore, not surprising that no relevant predictors could be identified. It is likely that the rarity of core-unexpressed elements (0.89% of cases) prevented stable pattern learning, raising questions about whether this constitutes a coherent coreness category.

Overall, the results indicate that FrameNet’s core-non-core distinction is robustly usage-based, while finer-grained coreness subtypes lack consistent distributional grounding, highlighting both the empirical strengths and theoretical limits of FrameNet’s current annotation system.

Several limitations suggest directions for future research. The analysis is restricted to verbal lexical units in English; extension to other parts of speech and languages

would test cross-linguistic validity. The extreme rarity of core-unexpressed elements warrants targeted investigation with expanded data collection or clearer annotation guidelines. The semantic content of influential word vector dimensions (particularly V163) remains unexplored and could reveal deeper connections between distributional semantics and argument structure. Finally, pursuing an unsupervised rather than supervised machine learning approach could reveal whether coreness distinctions naturally emerge from the feature space.<sup>15</sup>

## 4.2. Implications

### 4.2.1. For frame semantics and FrameNet

The present findings have non-negligible implications for frame semantics and the architecture of FrameNet. First, the strong performance of information-theoretic measures in distinguishing core from non-core elements validates the intuition in the literature that coreness fundamentally reflects argument predictability. This interpretation aligns with FrameNet's criterion that a core element 'always has to be overtly specified' (Ruppenhofer et al., 2016, p. 23), which translates empirically to high probability and low surprisal. However, the current analysis reveals that this probability-based criterion is less diagnostic than entropy deviation, that is, the predictability of an element relative to the frame's average. More broadly, these findings provide strong empirical support for the traditional argument/adjunct dichotomy, suggesting that while individual borderline cases exist, the core/non-core boundary reflects genuine distributional patterns.

The failure of the core versus core-unexpressed distinction raises more fundamental questions about FrameNet. Core-unexpressed elements, by definition, lack surface realisation and should not occur to begin with. The extreme instability in classification performance suggests that this category may not constitute a coherent distributional class but rather an idiosyncratic collection of annotations that are strongly frame-dependent, as was suggested by the elevated frame covariance parameter. The category's theoretical basis (elements 'absorbed by the lexical units in the frame'; Ruppenhofer et al., 2016, p. 25) may describe a semantic or lexical phenomenon that does not manifest in usage patterns. From a usage-based perspective, this disconnect between theoretical motivation and empirical realisation warrants reconsideration of whether core-unexpressed should be treated as a coreness category at all, potentially justifying their removal from FrameNet.

The peripheral versus extra-thematic distinction presents a different challenge. Usage data primarily improve the identification of peripheral elements, which exhibit distinct surprisal and word vector patterns. Entropy deviation is probably no longer helpful because non-core elements predominantly concentrate in the high deviation region, with peripheral ones being slightly less surprising than extra-thematic ones. What is striking is the importance of word vector dimension V163 across both core/non-core and peripheral/extra-thematic tasks, suggesting it may encode a general semantic property of participant roles or coreness, though further investigation of its content is needed, as word vectors are notoriously challenging to interpret (see Günther et al., 2019, for an in-depth discussion). In short, it appears likely that there

<sup>15</sup>I thank the anonymous reviewer who brought this to my attention.

are at least two subpopulations among the non-core elements, yet they are more difficult to delimit than in the core/non-core case.

#### 4.2.2. For cognitive research in argument realisation

The empirical validation of the core versus non-core distinction has direct consequences for research on argument null instantiation and for argument realisation more generally. Since null instantiation is conventionally restricted to core frame elements, establishing which elements qualify as core is foundational to defining the phenomenon's scope. Previous work has lacked an empirical basis for this determination, potentially including non-core elements in null instantiation analyses or excluding legitimate core elements. The information-theoretic and distributional semantic criteria examined in this study provide a principled and reproducible method for the identification of conceptually essential arguments across languages.

Given the features summarised in Section 4.1, it is possible to develop a coreness annotation pipeline without the need for computationally expensive classifiers. If, at the very least, frequency counts are available, an entropy-based decision rule could be used: An element  $e \in E_v^f$  belongs to the set of core elements  $C_v^f \subseteq E_v^f$  if and only if entropy deviation is less than or equal to zero (cf. Equation 13), that is,

$$C_v^f = \{e \in E_v^f : \Delta_v^f(e) \leq 0\}. \quad (13)$$

To validate this measure, the mixed-effects gradient boosting model was refitted with entropy deviation as a single predictor, yielding very good performance for both core (Precision = 0.89, Recall = 0.81, F1 = 0.85) and non-core outcomes (Precision = 0.82, Recall = 0.89, F1 = 0.85) with a normalised MCC of 0.85. As such, this very simple model is basically equivalent to the full, complex model reported in Table 8 earlier. Although this criterion may perform well as a practical decision rule, assuming a sharp boundary necessarily involves loss in nuance and further disregards the subtle influence of other distributional factors.

**Data availability statement.** The data that support the findings of this study are openly available in this OSF repository: <https://osf.io/n74e9>.

**Competing interests.** The author reports that there are no competing interests to declare.

## References

- Ágel, V. (2000). *Valenztheorie*. Narr.
- Alpaydm, E. (2022). *Maschinelles Lernen* (3rd rev. ed.). De Gruyter.
- Baayen, R. H. (2001). *Word frequency distributions*. Kluwer.
- Bird, S., Loper, E., & Klein, E. (2009). *Natural language processing with python*. O'Reilly Media Inc.
- Boas, H. C. (2020). A roadmap towards determining the universal status of semantic frames. In R. Enghels, B. Defrancq, & M. Jansegers (Eds.), *Empirical and methodological challenges* (pp. 21–52). De Gruyter. <https://doi.org/10.1515/9783110682588-002>.
- Boas, H. C., Ruppenhofer, J., & Baker, C. (2024). FrameNet at 25. *International Journal of Lexicography*, 37(3), 263–284. <https://doi.org/10.1093/ijl/eca009>.
- Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). *Enriching word vectors with subword information*. <https://arxiv.org/abs/1607.04606>
- Boleda, G. (2020). Distributional semantics and linguistic theory. *Annual Review of Linguistics*, 6, 213–234.

- Buskin, V. (2025). Definite null instantiation in English(es): A usage-based construction grammar account. *Constructions and Frames*, 17(2), 278–312. <https://doi.org/10.1075/cf.24007.bus>.
- Busse, D. (2012). *Frame-Semantik: Ein Kompendium*. De Gruyter.
- Chaves, R. P., Kay, P., & Michaelis, L. A. (2025). *Unrealized arguments and the grammar of context*. Cambridge University Press. <https://doi.org/10.1017/9781009663786>.
- Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, 21(1), 6. <https://doi.org/10.1186/s12864-019-6413-7>.
- Chicco, D., & Jurman, G. (2023). The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification. *Biodata Mining*, 16(1), 4. <https://doi.org/10.1186/s13040-023-00322-4>.
- Cover, T. M., & Thomas, J. A. (1991). *Elements of information theory*. John Wiley & Sons, Inc.
- Di Crescenzo, A., & Paolillo, L. (2021). Analysis and applications of the residual varentropy of random lifetimes. *Probability in the Engineering and Informational Sciences*, 35(3), 680–698. <https://doi.org/10.1017/S0269964820000133>.
- Fillmore, C. J. (1986). Pragmatically controlled zero anaphora. *Proceedings of the Twelfth Annual Meeting of the Berkeley Linguistics Society*, 12, 95–107. <https://doi.org/10.3765/bls.v12i0.1866>.
- Fillmore, C. J., Johnson, C. R., & Petruck, M. R. L. (2003). Background to FrameNet. *International Journal of Lexicography*, 16(3), 235–250.
- Fillmore, C. J., Lee-Goldman, R., & Rhomieux, R. (2012). The FrameNet construction. In H. C. Boas & I. A. Sag (Eds.), *Sign-based construction grammar* (pp. 309–372). CSLI Publications.
- Firth, J. R. (1968). *Selected papers of J. R. Firth, 1952–59*. Indiana University Press.
- Frank, S. L., Otten, L. J., Galli, G., & Vigliocco, G. (2015). The ERP response to the amount of information conveyed by words in sentences. *Brain and Language*, 140, 1–11. <https://doi.org/10.1016/j.bandl.2014.10.006>.
- Gavruseva, E. (2024). Let's drop the object: An experimental study of the implicit object construction in American English. *CogniTextes*, 25. <https://doi.org/10.4000/12rd2>.
- Givón, T. (2017). *The story of zero*. John Benjamins.
- Goldberg, A. E. (2001). Patient arguments of causative verbs can be omitted: The role of information structure in argument distribution. *Language Sciences*, 23(4), 503–524.
- Goldberg, A. E. (2005). Argument realization: The role of constructions, lexical semantics and discourse factors. In J.-Ö. Östman & M. Fried (Eds.), *Construction grammars: Cognitive grounding and theoretical extensions* (pp. 17–43). John Benjamins.
- Gries, S. T. (2020). Analyzing dispersion. In M. Paquot & S. T. Gries (Eds.), *A practical handbook of corpus linguistics* (pp. 99–118). Springer.
- Günther, F., Rinaldi, L., & Marelli, M. (2019). Vector-space models of semantic representation from a cognitive perspective: A discussion of common misconceptions. *Perspectives on Psychological Science*, 14(6), 1006–1033.
- Hamilton, W. L., Leskovec, J., & Jurafsky, D. (2016). Diachronic word embeddings reveal statistical laws of semantic change. In K. Erk & N. A. Smith (Eds.), *Proceedings of the 54th annual meeting of the association for computational linguistics*. <https://doi.org/10.18653/v1/P16-1141>
- Hastie, T., Tibshirani, R., & Friedman, J. H. (2017). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.
- Heidolph, K. E., Flämig, W., & Motsch, W. (1984). *Grundzüge einer deutschen Grammatik*. De Gruyter. <https://doi.org/10.1515/9783112642627>.
- Helbig, G. (1992). *Probleme der Valenz- und Kasus-theorie* (Vol. 51). Niemeyer.
- Herbst, T., & Götz-Votteler, K. (Eds.). (2007). *Valency: Theoretical, descriptive and cognitive issues*. De Gruyter.
- Herbst, T., Heath, D., Roe, I. F., & Götz, D. (2004). *A Valency dictionary of English: A corpus-based analysis of the complementation patterns of English verbs, nouns and adjectives*. Mouton de Gruyter.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning: With applications in R*. Springer.
- Janda, L. A. (Ed.). (2013). *Cognitive linguistics – The quantitative turn*. De Gruyter. <https://doi.org/10.1515/9783110335255>.

- Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science*, 20(2), 137–194. [https://doi.org/10.1207/s15516709cog2002\\_1](https://doi.org/10.1207/s15516709cog2002_1).
- Lenci, A. (2008). Distributional semantics in linguistic and cognitive research. *Rivista di Linguistica*, 20(1), 1–31.
- Levshina, N. (2019). Token-based typology and word order entropy: A study based on universal dependencies. *Linguistic Typology*, 23(3), 533–572. <https://doi.org/10.1515/lingty-2019-0025>.
- Linzen, T., & Jaeger, T. F. (2016). Uncertainty and expectation in sentence processing: Evidence from subcategorization distributions. *Cognitive Science*, 40(6), 1382–1411. <https://doi.org/10.1111/cogs.12274>.
- Liu, H. (2011). Quantitative properties of English verb valency. *Journal of Quantitative Linguistics*, 18, 207–233. <https://doi.org/10.1080/09296174.2011.581849>.
- Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. MIT Press.
- Marelli, M., & Baroni, M. (2015). Affixation in semantic space: Modeling morpheme meanings with compositional distributional semantics. *Psychological Review*, 122(3), 485–515.
- Matsubayashi, Y., Okazaki, N., & Tsujii, J. (2010). Generalization of semantic roles in automatic semantic role labeling. *Journal of Natural Language Processing*, 17, 59–89. [https://doi.org/10.5715/jnlp.17.4\\_59](https://doi.org/10.5715/jnlp.17.4_59).
- Mikolov, T., Grave, E., Bojanowski, P., Puhersch, C., & Joulin, A. (2018). Advances in pre-training distributed word representations. In N. Calzolari, K. Choukri, C. Cieri, T. Declerck, S. Goggi, K. Hasida, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, S. Piperidis & T. Tokunaga (Eds.), *Proceedings of the international conference on language resources and evaluation*. European Language Resources Association (ELRA).
- Mouselimis, L. (2024). *fast Text: Efficient learning of word representations and sentence classification using R* [R package version 1.0.4]. R package version 1.0.4. <https://CRAN.R-project.org/package=fastText>
- Nikolaev, D., & Padó, S. (2023). The argument–adjunct distinction in BERT: A FrameNet-based investigation. In M. Amblard & E. Breitholtz (Eds.), *Proceedings of the 15th international conference on computational semantics* (pp. 233–239). Association for Computational Linguistics. <https://aclanthology.org/2023.iwcs-1.23/>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Perek, F. (2016). Using distributional semantics to study syntactic productivity in diachrony: A case study. *Linguistics*, 54(1), 149–188. <https://doi.org/10.1515/ling-2015-0043>.
- Ross, S. M. (2010). *Introduction to probability models* (10th ed.). Academic Press.
- RStudio Team. (2020). *Rstudio: Integrated development environment for R*. RStudio, PBC. <http://www.rstudio.com/>
- Ruppenhofer, J., Ellsworth, M., Petruck, M. R. L., Johnson, C. R., Baker, C. F., & Scheffczyk, J. (2016). FrameNet II: Extended theory and practice. <https://framenet.icsi.berkeley.edu/wp/wp-content/uploads/2024/11/book.pdf>
- Ruppenhofer, J., & Michaelis, L. A. (2014). Frames and interpretation of omitted arguments in English. In S. K. Bourns & K. Lambrecht (Eds.), *Perspectives on linguistic structure and context: Studies in honor of Knud Lambrecht* (pp. 57–85). Benjamins.
- Say, S. (2025). Degrees of complexity in valency class systems: Implications for efficiency. *STUF - Language Typology and Universals*, 78(4), 607–661. <https://doi.org/10.1515/stuf-2025-2022>.
- Shafaei-Bajestan, E., Uhrig, P., & Baayen, R. H. (2022). Making sense of spoken plurals. *The Mental Lexicon*, 17(3), 337–367. <https://doi.org/10.1075/ml.22011.sha>.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(4), 623–656.
- Sigrist, F. (2022). Gaussian process boosting. *Journal of Machine Learning Research*, 23(232), 1–46.
- Slaats, S., & Martin, A. E. (2025). What’s surprising about surprisal. *Computational Brain & Behavior*, 8(2), 233–248. <https://doi.org/10.1007/s42113-025-00237-9>.
- Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, 128(3), 302–319. <https://doi.org/10.1016/j.cognition.2013.02.013>.
- Tagliamonte, S., & Baayen, H. (2012). Models, forests and trees of York English: WAS/were variation as a case study for statistical practice. *Language Variation and Change*, 24(2). <https://doi.org/10.1017/S0954394512000129>.

Tesnière, L., Osborne, T., & Kahane, S. (2015). *Elements of structural syntax*. John Benjamins Publishing Company.

Van Rossum, G., & Drake, F. L. (2009). *Python 3 reference manual*. CreateSpace.

Vater, H. (1978). On the possibility of distinguishing between complements and adjuncts. In W. Abraham (Ed.), *Valence, semantic case, and grammatical relations*. Benjamins.

Willems, R. M., Frank, S. L., Nijhof, A. D., Hagoort, P., & van den Bosch, A. (2016). Prediction during natural language comprehension. *Cerebral Cortex*, 26(6), 2506–2516. <https://doi.org/10.1093/cercor/bhv075>.

Willich, A. (2022). *Konstruktionsemantik*. De Gruyter. <https://doi.org/10.1515/9783110762341>.

Yoffe, G., Dershowitz, N., Vishne, A., & Sober, B. (2025). Estimating the influence of sequentially correlated literary properties in textual classification: A data-centric hypothesis-testing approach. *Journal of Quantitative Linguistics*, 1–33. <https://doi.org/10.1080/09296174.2025.2496172>.

### Appendix A Information theory

**Proposition A.1** (Property of entropy deviation). Let  $S_v^f(e)$  denote the surprisal of element  $e$  and  $H(E_v^f) = \mathbb{E}[S_v^f(e)]$  the entropy of its corresponding frame element distribution. Then, the entropy deviation  $\Delta_v^f(e)$  satisfies:

$$\mathbb{E}[\Delta_v^f(e)] = 0. \tag{14}$$

**Proof.**

$$\mathbb{E}[\Delta_v^f(e)] = \sum_e P_v^f(e) (S_v^f(e) - H(E_v^f)) \tag{15}$$

$$= \sum_e P_v^f(e) S_v^f(e) - H(E_v^f) \sum_e P_v^f(e) \tag{16}$$

$$= H(E_v^f) - H(E_v^f) \tag{17}$$

$$= 0. \tag{18}$$

**Proposition A.2** (Entropy variance). The expectation of the squared entropy deviation corresponds to the varentropy of the frame element distribution:

$$\mathbb{E}[(\Delta_v^f(e))^2] = \mathbb{V}[S_v^f(e)]. \tag{19}$$

**Proof.** The variance of a random variable  $X$  is defined as follows:

$$\mathbb{V}[X] := \mathbb{E}[X^2] - (\mathbb{E}[X])^2. \tag{20}$$

Therefore, the goal is to show that:

$$\mathbb{E}[(\Delta_v^f(e))^2] = \mathbb{E}[S_v^f(e)] - (\mathbb{E}[S_v^f(e)])^2. \tag{21}$$

After applying the definition of expectation (cf. Equation 7) to  $(\Delta_v^f(e))^2$ , expanding the square and summarising the remaining terms, the identity in Equation 21 follows naturally.

$$\mathbb{E}[(\Delta_v^f(e))^2] = \sum_e P_v^f(e) [S_v^f(e) - H(E_v^f)]^2 \tag{22}$$

$$= \sum_e P_v^f(e) [S_v^f(e)^2 - 2S_v^f(e)H(E_v^f) + H(E_v^f)^2] \tag{23}$$

$$= \sum_e P_v^f(e) S_v^f(e)^2 - \sum_e P_v^f(e) 2S_v^f H(E_v^f) + \sum_e P_v^f(e) H(E_v^f)^2 \tag{24}$$

$$= \sum_e P_v^f(e) S_v^f(e)^2 - 2H(E_v^f) H(E_v^f) + H(E_v^f)^2 \tag{25}$$

$$= \sum_e P_v^f(e) S_v^f(e)^2 - H(E_v^f)^2 \tag{26}$$

$$= \mathbb{E}[S_v^f(e)^2] - (\mathbb{E}[S_v^f(e)])^2 \text{ (definition of variance)} \tag{27}$$

$$= \mathbb{V}[S_v^f(e)] \tag{28}$$

### Appendix B Formal specification of word vector construction

This appendix provides a formal specification of the procedures used to construct distributional semantic representations of FrameNet frame elements.

#### B.1. Frame element realisation

For each occurrence  $i$  in which verb  $v$  evokes frame  $f$  (where  $i = 1, 2, \dots, N_v^f$ ), let  $\mathcal{W}_i^{(v,f)}(e)$  denote the ordered sequence of words that realise frame element  $e$  in that occurrence:

$$\mathcal{W}_i^{(v,f)}(e) = (w_1, w_2, \dots, w_{n_i}), \tag{29}$$

where  $n_i = |\mathcal{W}_i^{(v,f)}(e)|$  is the number of words in the realisation of element  $e$  in occurrence  $i$ . If element  $e$  is not present in occurrence  $i$ , then  $\mathcal{W}_i^{(v,f)}(e) = \emptyset$ .

#### B.2 Word embeddings

Let  $\varphi: W \rightarrow \mathbb{R}^d$  denote the word embedding function that maps each word  $w \in W$  (the vocabulary) to a  $d$ -dimensional vector representation, where  $d$  is the embedding dimensionality.

For each occurrence  $i$  of verb  $v$  evoking frame  $f$ , the sequence of word vectors corresponding to the realisation of frame element  $e$  is defined as follows:

$$\Phi_i^{(v,f)}(e) = (\varphi(w_1), \varphi(w_2), \dots, \varphi(w_{n_i})), \tag{30}$$

where  $w_j \in \mathcal{W}_i^{(v,f)}(e)$  for  $j \in \{1, 2, \dots, n_i\}$ . If  $\mathcal{W}_i^{(v,f)}(e) = \emptyset$ , then  $\Phi_i^{(v,f)}(e) = \emptyset$ .

#### B.3 Frame element vector representations

To obtain a single vector representation for each realised frame element, three cases are distinguished. If a frame element is not realised in a given occurrence, no vector representation is defined. If it is realised by a single word, the corresponding word embedding is used. If it is realised by multiple words, the unweighted arithmetic mean of their embeddings is computed.

Formally, the vector representation  $\mathbf{v}_i^{(v,f)}(e)$  of frame element  $e$  in occurrence  $i$  is defined as follows:

$$\mathbf{v}_i^{(v,f)}(e) = \begin{cases} \emptyset & \text{if } n_i = 0 \\ \varphi(w_1) & \text{if } n_i = 1 \\ \frac{1}{n_i} \sum_{j=1}^{n_i} \varphi(w_j) & \text{if } n_i > 1. \end{cases} \tag{31}$$

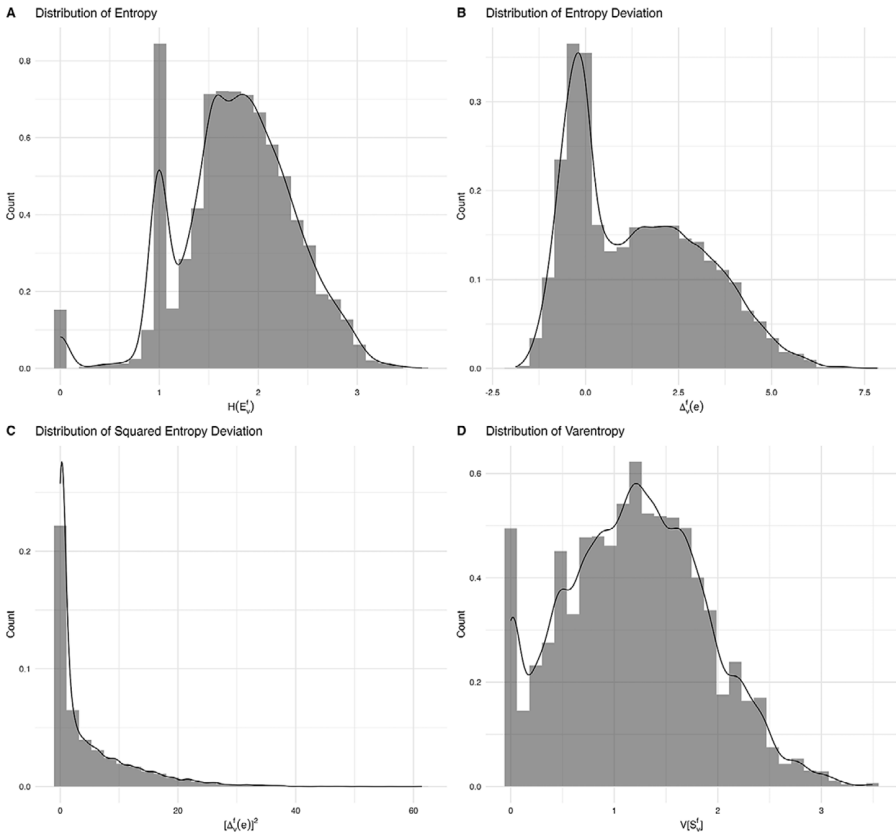
Instances for which  $\mathbf{v}_i^{(v,f)}(e) = \emptyset$  are excluded from subsequent aggregation.

Since the analysis operates at the level of verb–frame–element combinations, token-level frame element vectors are further aggregated across all attestations. Let  $O_v^f(e)$  denote the observed frequency of element  $e$  with verb  $v$  in frame  $f$ . The aggregated vector representation of frame element  $e$  is then given by the following:

$$\mathbf{v}_v^f(e) = \frac{1}{O_v^f(e)} \sum_{i:e \text{ attested in } i} \mathbf{v}_i^{(v,f)}(e). \tag{32}$$

These aggregated vectors, which represent centroids of all token-level vectors in embedding space, constitute the distributional semantic representations used in the main analysis.

### Appendix C Supplementary Figures



**Figure C1.** Unconditional distributions of information-theoretic measures (density plots). A: Distribution of entropy. B: Distribution of entropy deviation. C: Distribution of squared entropy deviation. D: Distribution of varentropy.

**Cite this article:** Buskin, V. (2026). Is FrameNet usage-based? Predicting frame element coreness with information theory and gradient boosting, *Language and Cognition*, 18, e30, 1–31. <https://doi.org/10.1017/langcog.2026.10076>